FAST 工程主动反射面健康监测系统中温度信号多元回归估测*

孙　晓[1]，王清梅[2]，李振伟[1]，乔峰[1]，楚敬敬[1]

（1.青岛科技大学 自动化与电子工程学院，山东青岛 266061；2.中国科学院国家天文台，北京100012）

摘 要：大型结构健康监测系统中，温度测点故障会造成安全隐患。基于FAST工程主动反射面健康监测系统，提取并分析了9个测点数据间的线性相关性，将候选变量分组并选择最优回归子集，建立多元线性回归模型，融合正常测点数据对故障测点进行估测。针对变量间多重共线性问题进一步应用了岭回归法，选定岭参数为6。使用F检验与拟合度检验了模型显著性和有效性，使用不同时日数据验证了模型的估测精度。结果证明，多元线性回归模型比单变量具有更高的拟合度与精度，均方根误差为0.475℃，岭回归法稳定性更高，均方根误差仅为0.435℃。

关键词：500米口径球面射电望远镜；温度传感器；结构健康监测；多元线性回归；岭回归

中图分类号：TP277;TP111.44　　**文献标志码：**A　　**文章编号：**

0 引 言

结构健康监测是通过布设大量传感器，监测并评估结构当前状态的一种技术，在工程结构的安全监测中有着广泛应用。500米口径球面射电望远镜FAST(Five- hundred-meter Aperture Spherical Telescope)[1]大型射电望远镜为口径500米具有主动反射面的大型钢结构工程。文献[2]讨论了FAST工程结构受力复杂性，由于其结构特殊、空间跨度大且环节众多，在观测过程中主动反射面受控变形致结构受力复杂，为保证其安全工作，建设了主动反射面健康监测系统，通过监测关键结构的应力与环境信息，实现FAST望远镜的健康状态评估。文献[3]指出，在索网施工过程中，该系统所监测圈梁实时应力最大约60Mpa，始终小于设计安全值201.5Mpa，保持在安全范围内，并指出了温度信号对应力分析的重要性。

在FAST运行期间，工作过程中索网主动变形会使结构受力更为复杂，以2019年7月份监测数据为例，所有圈梁格构柱测点应力值最大值为137.37Mpa，最小值为-135.88Mpa，仍处于安全范围内，但部分测点应力变化范围较大，应力变化最大的测点其极差为167.17Mpa。在望远镜后续工作中，应力的监测对结构实时健康状态评估、长期疲劳损伤评估等仍具有重要意义。

大型工程结构空间尺寸大，不同部位温升不同，温差与温度效应会使结构应力发生较大变化，文献[4]指出FAST工程结构中，温度荷载对圈梁结构的刚度起控制作用，分布测点温度信息是评估结构状态的重要依据[5]。且在FAST工程中，应用了416只光纤Bragg光栅(Fiber Bragg Grating, FBG)应变传感器[6]监测应力，共包含圈梁及格构柱应力测点100只，主索索力测点316只，虽测点分布位置不同，但所有测点均需根据不同测点的实际温度补偿应变传感器的温度应变交叉敏感问题，文献[7]进一步讨论了分离结构应力的温度效应对状态评估的必要性，温度信息缺失将直接导致测点应变信息不可靠。

实际工程现场环境复杂，大量传感器长期工作，传感器及数据通道存在一定故障率。检修发现，FAST主动反射面健康监测系统已有个别光纤光栅解调仪出现故障，且存在数只传感器数据异常，已确定异常测点数目占比约为5%，使部分监测数据无法正常获取。但受限于现场施工条件，以及部分测点安装位置特殊，故障传感器常无法第一时间修复，数据的缺失将导致健康监测系统性能下降，形成安全隐患。

使用临近测点值代替故障测点温度信息，就FBG传感器而言，因其常串联布设，若数据

***基金项目：**国家自然科学基金资助项目(11803053)资助
收稿日期：2021-11-02；　**修订日期：**2021-11-19
作者简介：孙晓，男，讲师，博士.研究方向：结构健康监测、天文技术与方法.Email:sx@qust.edu.cn

通道发生损坏，会使临近的一组传感器均失效。使用较远传感器监测值代替，存在距离过远，光照影响温差较大，代替不准确等问题，这也是每测点都布设温度传感器的原因。实现故障测点的温度信号准确估测，对提高健康监测系统可靠性有着重要意义。

1　传感器布设与数据提取分析

FAST主动反射面结构健康监测系统圈梁格构柱关键应变测点的温度信息，使用FBG温度传感器监测。传感器工作原理是外界温度引起光栅周期以及有效折射率发生变化，使反射波长偏移，通过测量中心波长变化量，获取传感器所测温度。

为有效探究远距离布设的温度测点之间的相关性，提取FAST边缘圈梁支承格构柱对应测点的温度信息，每个被监测的格构柱提取一路测点信息，FAST工程圈梁格构柱共50个，有10个格构柱装有测点，被测格构柱编号与测点分布如图1所示，传感器安装于如图2所示圈梁支座处的靠内侧的水平拉杆中的绿色杆件，测点之间最远距离为500m。

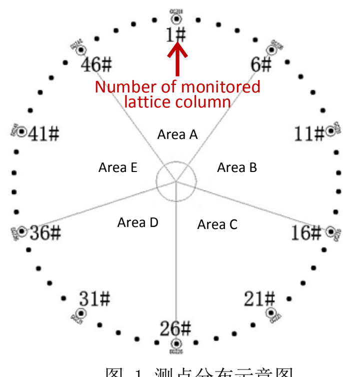


图 1 测点分布示意图

Fig.1 Distribution map of measuring points

图 2　传感器安装位置

Fig.2 Sensor mounting position

监测系统数据采集频率为1Hz，为降低数据量，滤除高频干扰，且避免由于数据长度不统一造成的分析困难。对原始数据进行平均处理，平均时窗为10min，取每10分钟内温度数据平均值作为该时刻温度数据，每测点每日得144个数据样本。提取2019年7月份监测数据，对应记作$T_{1\#} \sim T_{46\#}$，其中31#格构柱测点因采集通道检修关闭无数据记录，共提取9个测点信息。因环境温度循环周期往往为昼夜，按天对数据进行分组，7月16日数据如图3所示。

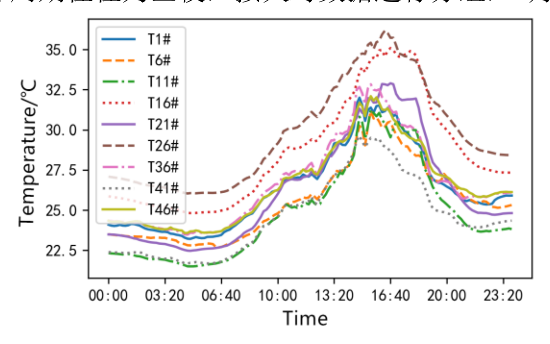


图 3 温度监测数据

Fig.3 Temperature monitoring data

可见不同测点之间温差较大，最大温差超过5℃，波动规律也有一定区别，但其趋势大体一致。两数据样本分别记作$X=[x_1,x_2,...,x_n]$，$Y=[y_1,y_2,...,y_n]$，计算不同样本间线性相关性，相关系数公式为：

$$r_{x,y}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

式中：n为样本长度，$\bar{x},\bar{y}$为各自均值，得不同测点温度数据相关系数矩阵如表1，可见不同测点之间存在高度线性相关，相关系数最低为0.959。

表 1 不同测点温度数据相关系数矩阵表

Tab.1 Correlation coefficient matrix of temperature data

	$T_{1\#}$	$T_{6\#}$	$T_{11\#}$	$T_{16\#}$	$T_{21\#}$	$T_{26\#}$	$T_{36\#}$	$T_{41\#}$	$T_{46\#}$
$T_{1\#}$	1.000	0.973	0.986	0.969	0.982	0.984	0.997	0.996	0.996
$T_{6\#}$	0.973	1.000	0.980	0.981	0.974	0.984	0.967	0.969	0.982
$T_{11\#}$	0.986	0.980	1.000	0.991	0.993	0.988	0.981	0.975	0.985
$T_{16\#}$	0.969	0.981	0.991	1.000	0.994	0.992	0.959	0.959	0.972
$T_{21\#}$	0.982	0.974	0.993	0.994	1.000	0.995	0.974	0.973	0.981
$T_{26\#}$	0.984	0.984	0.988	0.992	0.995	1.000	0.977	0.980	0.988
$T_{36\#}$	0.997	0.967	0.981	0.959	0.974	0.977	1.000	0.996	0.995
$T_{41\#}$	0.996	0.969	0.975	0.959	0.973	0.980	0.996	1.000	0.997
$T_{46\#}$	0.996	0.982	0.985	0.972	0.981	0.988	0.995	0.997	1.000

2 线性回归模型

2.1 模型简介

当变量之间高度线性相关时，常用线性回归方法建模，自变量个数为1时，记为x，因变量记作y，其一般模型记为：

$$y=\beta_0+\beta_1 x+\varepsilon \qquad (2)$$

式中：β_0，β_1为模型系数，ε为随机误差。

健康监测温度估测应用中，受光照不均匀的影响，一元回归在某些时刻预测精度较差[8]。引入更多变量，综合不同测点的温度影响，可取得更优的预测结果。当变量个数为多个x_1，x_2,…，x_p，称为多元线性回归[9]，模型记为：

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p+\varepsilon \qquad (3)$$

式中：β_0，β_1,…，β_p为系数，p为变量个数，当获得n组数据时，模型可简记为矩阵形式：

$$y=XB+E \qquad (4)$$

$$Y=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} B=\begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} E=\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \qquad (5)$$

$$X=\begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix} \qquad (6)$$

式中：Y为因变量向量，X为自变量矩阵，B为各自变量系数向量，E为随机误差向量，随机误差项符合正态分布$\varepsilon_i \sim N(0,\sigma^2), i=1,2,\ldots,n$。舍去误差项的影响，最小二乘法求解模型的自变量系数，系数向量估计为：

$$\hat{B}=(X^T X)^{-1} X^T Y \qquad (7)$$

利用所得自变量估计系数值$\hat{\beta}_0,\hat{\beta}_1,\ldots,\hat{\beta}_p$即可建立经验回归方程：

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1 x_1+\ldots+\hat{\beta}_p x_p \qquad (8)$$

式中： $x_1 \sim x_p$ 为自变量输入值， $\hat{y}$ 为模型输出估测值。

对提取的9个测点数据，取其一测点作为因变量即被估测点，其余作为自变量输入，即可通过训练建立多元线性回归模型，当被估测点数据缺失时，可利用自变量输入，对其信号进行估测。

2.2　最优自变量选取

将1#格构柱测点作为被估测点，其余8个测点温度信息作为候选变量，其模型变量的不同分组子集有 2^8-1 种。利用16日数据训练模型，将变量按照相关性降序排列逐个增加，划分为8组。

结合 C_p 统计量和 R_a^2 自由度调整复决定系数两选取准则来确定最优变量的子集：

$$C_p=(n-m-1)\frac{RSS_p}{RSS_m}-n+2p \qquad (9)$$

式中：n=144为样本长度，m=8为全部待选变量个数，p为模型所选变量个数， RSS_m 为选择全部变量建模的残差平方和， RSS_p 为模型残差平方和。

$$R_a^2=1-\frac{n-1}{n-p-1}(1-R^2) \qquad (10)$$

式中： R^2 为复决定系数即模型拟合度：

$$R^2=1-\frac{RSS}{TSS} \qquad (11)$$

式中：TSS为因变量y=$T_{1\#}$ 离差平方和。

$$RSS=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \qquad (12)$$

$$TSS=\sum_{i=1}^{n}(y_i-\bar{y})^2 \qquad (13)$$

式中： y_i 为实际值， $\hat{y}_i$ 为模型预测值， $\bar{y}$ 为实际值均值。

变量分组与对应模型 C_p 和 R_a^2 数值如表2所示，选取 C_p 最小且 R_a^2 最大，即为最优变量子集，可见八元作为变量时模型拟合度最优。

表 2　变量分组与模型 C_p、 R_a^2 数值

Tab. 2 Variable grouping and C_p, R_a^2 value of model

Groups	Variables	C_p	R_a^2
Univariate	$T_{36\#}$	130.51	0.9937
Two variables	$T_{36\#}$,$T_{41\#}$	98.498	0.9944
Three variables	$T_{36\#}$,$T_{41\#}$,$T_{46\#}$	56.04	0.9954
Four variables	$T_{11\#}$,$T_{36\#}$,$T_{41\#}$,$T_{46\#}$	20.95	0.9963
Five variables	$T_{11\#}$,$T_{26\#}$,$T_{36\#}$,$T_{41\#}$,$T_{46\#}$	22.01	0.9963
Six variables	$T_{11\#}$,$T_{21\#}$,$T_{26\#}$,$T_{36\#}$,$T_{41\#}$,$T_{46\#}$	22.01	0.9963
Seven variables	$T_{6\#}$,$T_{11\#}$,$T_{21\#}$,$T_{26\#}$,$T_{36\#}$,$T_{41\#}$,$T_{46\#}$	13.29	0.9965
Eight variables	$T_{6\#}$,T1$_{1\#}$,$T_{16\#}$,$T_{21\#}$,$T_{26\#}$,$T_{36\#}$,$T_{41\#}$,$T_{46\#}$	7	0.9967

3　多重共线性问题与岭回归模型

由表1知，所有测点两两之间均存在极强相关性，对于多元线性回归模型，会因多重共线性使模型稳定性出现问题，即对模型可能存在一组数 $k_0,k_1,\ldots,k_p$ 使：

$$k_0 + k_1 x_{1i} + \ldots + k_p x_{pi} \approx 0, (i = 1,2,\ldots,n) \qquad (14)$$

此时设计矩阵秩 $rank(X) < p+1$，意味着公式(7)中 $|X^T X| \approx 0$，而模型系数向量估计矩阵 $\hat{B}$ 方差矩阵 $D(\hat{B}) = \sigma^2 (X^T X)^{-1}$，此时方差阵对角元素过大，使自变量系数向量B的估计精度下降，模型对输入数据的细小变化会过于敏感，预测结果不确定性增加，模型的稳定性和精确度降低。

岭回归法针对多元线性回归模型中的多重共线性问题进行了优化[10]，向量系数岭估计定义为：

$$\hat{B}(\lambda) = (X^T X + \lambda I)^{-1} X^T Y \qquad (15)$$

式中：$\lambda > 0$ 称为岭参数。可见给 $X^T X$ 增加了一个正常数矩阵L2范数惩罚项，保证 $X^T X + \lambda I$ 满秩可逆，接近奇异程度比原矩阵减小。$X^T X + \lambda I$ 随着岭参数增大而增大，模型方差减小，但向量系数估计值 $\hat{B}$ 将偏离原值，使模型偏差增大，$\lambda = 0$ 模型退化为普通多元线性回归模型，在应用时，需合理选取岭参数。

应用岭迹法，将8个自变量回归系数对应岭迹绘制如图4所示，按原则选取使所有向量系数稳定的最小的一个岭参数值，确定 $\lambda = 6$，建立岭回归模型。由图4可见，变量 $T_{6\#}$、$T_{26\#}$ 系数趋近于0，说明岭回归法训练期间，对自变量有筛选作用。

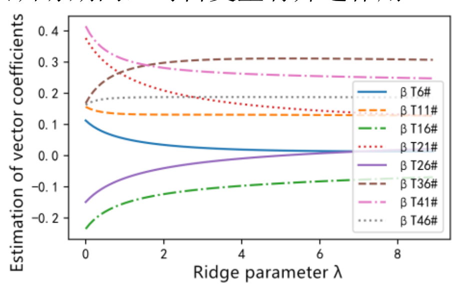


图 4 岭迹图

Fig.4 Ridge parameter map

4 模型检验与评价

4.1 检验与评价指标

多元线性回归模型建立后，需判断模型是否可靠，通过F检验确定模型建立的回归方程显著性，统计量公式为：

$$F = \frac{ESS/p}{RSS/(n-p-1)} \qquad (16)$$

式中：RSS为模型残差平方和，ESS为回归平方和，n=144为样本长度，p=8为所选变量个数。

$$ESS = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad (17)$$

式中：$\hat{y}_i$ 为模型预测值，$\bar{y}$ 为实际值均值。

模型拟合度 R^2 检验，见公式(11)，值越接近于1说明模型拟合度越高。

以$T_{1\#}$数据作为被估测点，其他测点数据作为输入变量。使用16日数据训练建立模型，并对17日、18日数据进行估测检验，使用均方根误差RMSE评价模型的估测效果，其值越小说明模型预测效果越好：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad (18)$$

4.2 测试结果

对照前述分组，建立从一元到八元不同类型的线性回归与岭回归模型，并对模型进行估测效果评价，检验及模型估测结果见表3。F检验统计量最低为4972.29，依旧远大于查表所得$F_a(1,135)$= 3.911，所有模型通过显著性检验。拟合度均在0.99以上，满足拟合度要求。随着融合变量的增加，模型测试的RMSE也随之减小。

表 3 模型检验与应用测试结果

Tab. 3 Model test and application results

Groups	*F* test	16th model fitting R^2 Fitting degree	16th model fitting *RMSE*/℃	17th estimate *RMSE*/℃	18th estimate *RMSE*/℃	Estimated mean value *RMSE*/℃
Univariate	22517.92	0.9937	0.2	0.68	0.66	0.67
Two variables	12784.23	0.9945	0.19	0.71	0.56	0.635
Three variables	10411.94	0.9955	0.17	0.52	0.46	0.49
Four variables	9593.81	0.9964	0.15	0.52	0.44	0.48
Five variables	7665.63	0.9964	0.15	0.52	0.45	0.485
Six variables	6878.33	0.9967	0.15	0.53	0.43	0.48
Seven variables	5876.86	0.9967	0.15	0.53	0.43	0.48
Eight variables	5418.98	0.9969	0.14	0.52	0.43	0.475
Ridge regression	4972.29	0.9966	0.15	0.46	0.41	0.435

其中，一元、八元线性回归、岭回归模型的预测效果可见如图5所示，预测曲线与真实值几乎重叠，几种模型均可较好的跟踪估测信号，岭回归法模型精度更高。

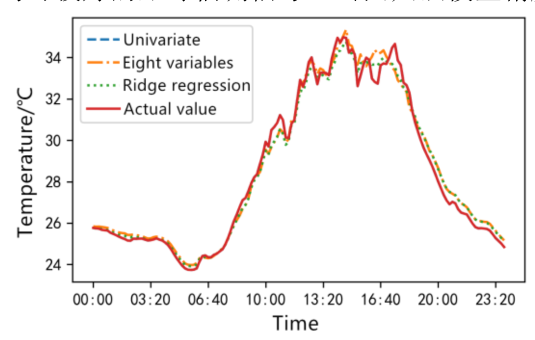


图 5 真实值与模型预测值对比

Fig. 5 Comparison of true value and model predicted value

5 结 语

提取FAST工程主动反射面健康监测系统中跨度500m范围布设的9个测点温度监测数据，分析知两两间存在极强的线性相关性，将1#测点作为被估测点，研究了融合其余测点数据对其估测的方法，并对比了不同变量集时模型的估测效果。

多元线性回归对比一元的回归模型，无论是在拟合度还是实际估测应用中，都具有更好的效果，当输入变量增加至三元时，测试中RMSE<0.5℃。而岭回归对比原始多元线性回归，模型拟合度略有降低，但可有效避免变量间的多重共线性问题，对比17、18日数据的测试，其预测精度要更优于多元回归，算法具有更强的抗干扰能力，稳定性好，两日平均RMSE仅为0.435℃。

结果表明，当某测点出现故障时，可使用岭回归法利用近期历史数据训练并建立多元线性回归模型，将正常测点温度信息作为输入，对故障测点输出值进行估测。方法具有较高精度，可用于补充健康监测系统缺失的节点温度信息，在故障测点维修期间保持健康监测系统的正常工作。但该方法使用近期日期数据进行拟合预测，只能作为测点故障后短期的替补方案，若测点故障过久，可用数据时日相差较大，由于日照天气差别等原因，设备温度场变化规律迁移，估测精度将较低，长期数据缺失的估测方法仍需进一步研究。

参考文献

[1]NAN R D. Five Hundred Meter Aperture Spherical Radio Telescope (FAST)[J]. Science in China: Series G:Physics Mechanics

&Astronomy,2006,49(2):129-148.
[2]王晓琳, 李东平, 朱明, 等. FAST主动反射面健康监测系统设计与应用研究[J]. 广西科技大学学报,2018,29(4):84-91.
WANG X L, LI D P, ZHU M, et al. Design and application of health monitoring system of main reflector system for FAST[J]. Journal of Guangxi university of Science and Technology, 2018,29(4): 84-91.
[3]王玉婷, 王启明, 朱明, 等. 光纤光栅传感器在500m口径球面射电望远镜工程索网施工阶段的应用[J]. 中国机械工程, 2016,27(20):2760-2764.
WANG Y T, WANG Q M, ZHU M, et al.Application of Fiber Grating Sensor in Construction Stage of Cable Net for FAST[J]. China Mechanical Engineering, 2016,27(20):2760-2764.
[4]赵保庆, 王启明, 李志恒, 等. FAST圈梁支承结构性能理论与实验研究[J]. 工程力学, 2018,35(S1):200-211.
ZHAO B Q, WANG Q M, LI Z H, et al. Performance research of FAST ring beam by theoretical and experimental[J]. Engineering Mechanics, 2018,35(S1):200-211.
[5]王惠,宁云炜, 闫浩. 26米天线座架温度分布规律及变形影响分析[J]. 天文研究与技术, 2018,15(02): 208-215.
WANG H, NING Y W, YAN H. Temperature Distribution and Deformation Impact Analysis of 26m Antenna Frame[J]. Astronomical research and technology, 2018,15(02):208-215.
[6]孙晓, 王启明, 吴明长, 等. 光纤Bragg光栅应变计在500米口径球面射电望远镜工程索力监测中的应用[J]. 光学精密工程, 2015, 23(4):919-925.
SUN X, WANG Q M, ZHU M et al. Application of optical fiber Bragg grating strain gauge to cable force monitoring of FAST[J]. Optics and Precision Enginerring,2015. 23 (4) : 919-925.
[7]王清梅, 朱明, 王启明, 等. FAST主动反射面健康监测系统数据处理方法研究及应用[J].天文研究与技术, 2017,14(2):164-171.
WANG Q M, ZHU M, WANG Q M, et al. Research and application of Data Processing Method of the Health Monitoring System for the Main Active Reflector of FAST[J]. Astronomical research and technology, 2017,14(2): 164-171.
[8]郭忠臣, 孙朋, 李致春, 等. 预报模型及建模序列长度对钟差短期预报精度影响研究[J]. 天文研究与技术, 2020,17(03):299-307.
GUO Z C, SUN P, LI Z C, et al. Research on the Influence of Prediction Model and Modeling Sequence Length on Short-term Prediction Accuracy of Clock Bias[J]. Astronomical research and technology, 2020,17(03):299-307.
[9]肖霞, 伍兴国. 线性回归中多重共线性的几何解释[J]. 统计与决策, 2021(21):46-51.
XIAO X, WU X G. Geometric Interpolation of Multicollinearity in Linear Regression[J]. Statistics & Decision, 2021(21):46-51.
[10]HOERL A E, KENNARD R W. Ridge regression: application for non-orthogonal problems[J]. Techometrics, 1970(12):69-72.

Multi-linear regression model to estimate temperature signal in FAST project active reflective surface health monitoring system

Sun Xiao[1], Wang Qingmei[2], Li Zhenwei[1],Qiao Feng[1], Chu Jingjing[1]
(1. College of Automation and Electronic Engineering, Qingdao University of Science and Technology, Qingdao 266061, China;2.National Astronomical Observatories, Chinese Academy of Science, Beijing 100012, China)

Abstract: The temperature measurement point malfunction in the large-scale structural health monitoring system could cause safety hazards. The paper extracts 9 temperature measurement points' data of the FAST project active reflective surface health monitoring system and analyzes the linear correlation of them. In order to establish a multiple linear regression model, the variables are grouped and the optimal subset is selected. The fault measuring point is estimated by fusing the normal measuring points' data. Focusing to the multiple collinear issue between variables, the ridge regression method is applied and the ridge parameter is set to 6. Using F-test and fit-degree, the significance and validity of models are tested. Based on different time data, the estimation accuracy is verified. The results show the multiple linear regression model has a higher degree of fit and accuracy than one variable model, considering its RMSE is 0.475℃. Besides, the ridge regression model is more stable, considering its RMSE is 0.435℃.

Keywords: Five-hundred-meter Aperture Spherical Telescope(FAST); temperature sensor; health monitoring system; multiple linear regression; ridge regression